Acoustic Metadata Management and
Transparent Access to Networked Oceanographic Data Sets

Marie A. Roch
Dept. of Computer Science, San Diego State University
5500 Campanile Drive, San Diego, CA 92182-7720
Phone: (619) 594-5830 FAX: (619) 594-6746 E-mail: marie.roch@sdsu.edu

Simone Baumann-Pickering
Scripps Institution of Oceanography
University of California at San Diego
9500 Gilman Dr., La Jolla, CA 92093-0205
Phone: (858) 534-7280 FAX: (858) 534-6849 E-mail: sbaumann@ucsd.edu

Catherine L. Berchok
Alaska Fisheries Science Center, National Marine Fisheries Service
National Oceanic and Atmospheric Administration, Department of Commerce
7600 Sand Point Way N.E., Building 4
Seattle, Washington 98115-6349
Phone: (206) 526-6331 FAX: (206) 526-6615 E-mail: Catherine.Berchok@noaa.gov

John A. Hildebrand
Scripps Institution of Oceanography
University of California at San Diego
9500 Gilman Dr., La Jolla, CA 92093-0205
Phone: (858) 534-4069 FAX: (858) 534-6849 E-mail: jhildebrand@ucsd.edu

Erin M. Oleson
Pacific Islands Fisheries Science Center, National Marine Fisheries Service
National Oceanic and Atmospheric Administration, Department of Commerce
1601 Kapiolani Blvd, Ste. 1110
Honolulu, Hawaii 996814
Phone: (808) 944-2172 FAX: (808) 941-0307 E-mail: Erin.Oleson@noaa.gov

Sofie M. Van Parijs
Northeast Fisheries Science Center, National Marine Fisheries Service
National Oceanic and Atmospheric Administration, Department of Commerce
166 Water Street, Woods Hole, MA 02543
Phone: (508) 495-2119 FAX: (508) 495-2258 E-mail: sofie.vanparijs@noaa.gov

Award Number: *N00014-11-1-0697*
http://cetus.ucsd.edu

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Acoustic Metadata Management and Transparent Access to Networked Oceanographic Data Sets**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **San Diego State University ,Department of Computer Science,5500 Campanile Drive,San Diego,CA,92182-7720**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

LONG-TERM GOALS

The long-term goals of this effort are to produce software capable of organizing and archiving metadata associated with the detection of marine mammals. The focus is on acoustic detections, but other modalities of detection as well as measurements such as conductivity, temperature, and depth (CTD) casts are included. In addition, the software provides interfaces to access oceanographic measurements from other data repositories in a transparent manner. Data shall be accessible from a variety of languages used by the scientific community for analysis and modeling.

OBJECTIVES

The objectives of this effort are to produce:

1. A database which can flexibly store multiple types of metadata derived from a variety of acoustic platforms, both stationary and mobile.
2. Standardization of methods to make the data repositories useful to the passive acoustic monitoring community.
3. Secure access on network platforms using industry standard security protocols.
4. Visualization primitives in selected analysis and modeling languages (e.g. Matlab, R).
5. Access methods for the above languages.
6. Primitives to query data both spatially and temporally in an efficient manner.
7. Demonstration projects to show the value of the database as a scientific workbench component.

APPROACH AND WORK PLAN

1) Technical approach
The acoustic metadata database enables researchers to organize, store and most importantly query information derived from passive acoustic monitoring (PAM). Due to the large number of acquisition platforms, types of detection effort, etc., this is a complicated semi-structured task and traditional databases do not meet the needs of PAM users. By compiling a large team of PAM users who work on a global scale, we are defining data standards that are likely to meet the needs of the PAM community in general. Networking capabilities provide the ability to share data and eventually export summary data to OBIS-SEAMAP. In addition, this effort provides users with access to online physical oceanography databases using a single interface. The project also provides access methods for a variety of computer languages used for analysis by the scientific community.

2) Key Personnel
Dr. Marie A. Roch (San Diego State University) is the project manager and administrator for this project. She also takes lead for the software development.
Dr. John A Hildebrand (Scripps Institution of Oceanography (SIO)) is the project manager for the subaward to SIO.
Drs. Simone Baumann-Pickering (Scripps Institution of Oceanography),
Catherine L. Berchok (NOAA Alaska Fisheries Science Center (AFSC)),
Erin M. Oleson (NOAA Pacific Island Fisheries Science Center (PIFSC)),
Melissa Soldevilla (NOAA Southeast Fisheries Science Center (SEFSC)), and
Sofie Van Parijs (NOAA Northeast Fisheries Science Center (NEFSC)), all represent data providers who will be using the database and are integrally involved in the operational specification, requirements, and testing.

Dr. Simone Baumann-Pickering is providing the lead on habitat modeling, and Dr. Sofie Van Parijs is the project manager for NOAA as well as the lead on data standardization.

3) Work plans for the upcoming year

Plans for the coming year include the initial public release of the software (Spring 2012), an internal workshop in February 2012, definitions of XML schema, and further work on mediators, user and data level security. We will also begin to design indices to efficiently support temporal-spatial queries. Associated personnel will also be using the database to complete PAM studies.

WORK COMPLETED

Work was scheduled to start in July 2011, but delays in contracting resulted in the bulk of the funds being inaccessible until October 2011 and the project is just getting started. Dr. Michael Weise at ONR is aware of the situation and we are working with him to adjust the funding schedule to reflect this delay. Progress completed to date can be placed in several categories: standardization, access to oceanographic datasets through the database, and data analysis.

Our standardization work consists of specifying methods for data preparation prior to intake into the database and how it can be presented subsequently using standards such as ISO 19115 or SensorML from the Open Geospatial Consortium. Discussions on data preparation have been fruitful and we have decided on data coding system, representation of effort, etc., and are now beginning to examine modifications to have automated cetacean call detectors and classifiers generate appropriate outputs for intake into the database. We have also engaged in discussions with Dan Kowal's group at the National Geographic Data Center (NGDC) on standards for metadata from instruments. Anna Milan has been identified as our liaison with NGDC in these efforts. We have identified areas where the current standards are inappropriate and will work with NGDC to establish appropriate extensions.

Access to oceanographic data is implemented via mediator services which translate database requests into the appropriate set of commands to retrieve data from various providers. The prototype mediator for ephemeris data has been extended, permitting the retrieval of lunar illumination and phase. We have also begun working on other mediators. We are building upon the work of Jason Roberts in Pat Halpin's lab, integrating the Marine Geospatial Ecology (GeoEco) toolset into our database services. While there is a steep learning curve as GeoEco is an evolving project with documentation in flux, Halpin's lab has been helpful and thanks to their assistance we have met our first milestone of connectivity to another data provider despite our late start.

Finally, we have begun analyzing data in preparation for habitat modeling, the results of which were presented at the 162nd meeting of the Acoustical Society of America (Baumann-Pickering et al. 2011). A brief summary of these preliminary results from this study are presented below. The project was significantly accelerated by use of the metadata database's ability to access external data sources.

RESULTS

Technical accomplishments include extending the mediator services to retrieve lunar illumination and phase, developing standards within the team on data preparation, and implementing first stage connectivity, the ability to connect and retrieve data, to data providers such as the National Data Buoy Center.

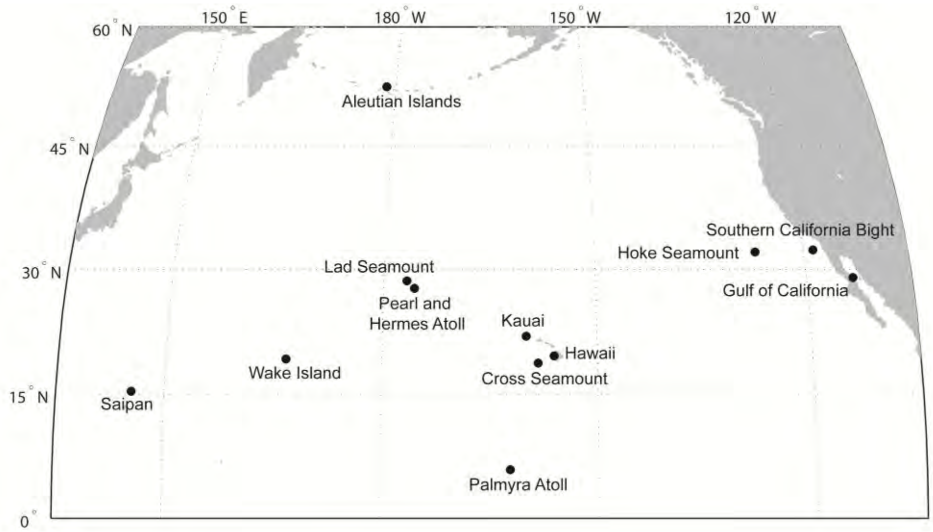

Figure 1 – Locations of high frequency acoustic recording packages (HARPs) throughout the North Pacific and Gulf of California between 2005 and 2010.

A preliminary study was done utilizing 100 kHz bandwidth 16-bit data from high frequency acoustic recording packages (HARPs; Wiggins and Hildebrand 2007) deployed by Scripps Institution of Oceanography and NOAA PIFSC throughout the North Pacific and Gulf of California (Fig. 1). A total of 35 deployments were analyzed from 12 sites. Deployment depths and mission lengths varied, with data coverage between the years 2005 and 2010. Coverage varied by site with some sites such as the Southern California Bight having extensive coverage with continuous recording and more remote sites such as Saipan having fewer deployments and duty cycled data. The median deployment depth was approximately 700 m, with outliers consisting of two shallow deployments in approximately 120 m of water at Lad Seamount. Other deployments at this site were at more typical depths. HARPs were not present at all sites at all times, preventing seasonal analysis of the entire dataset at this time.

In a number of locations, a low frequency acoustic diel pattern was apparent, occurring in a 1-6 kHz band (Fig. 2). A second low frequency diel pattern appeared at Wake Island only at 6-8 kHz (Fig. 2). Additionally, a broadband diel pattern was observed at some locations, ranging from very low frequencies up to the analysis cutoff frequency of 95 kHz. To further explore this pattern, ambient noise was estimated in two frequency bands, 1-6 kHz and 6-95 kHz, during the quietest 75 s (a convenient unit for data associated with HARPs) of each hour.

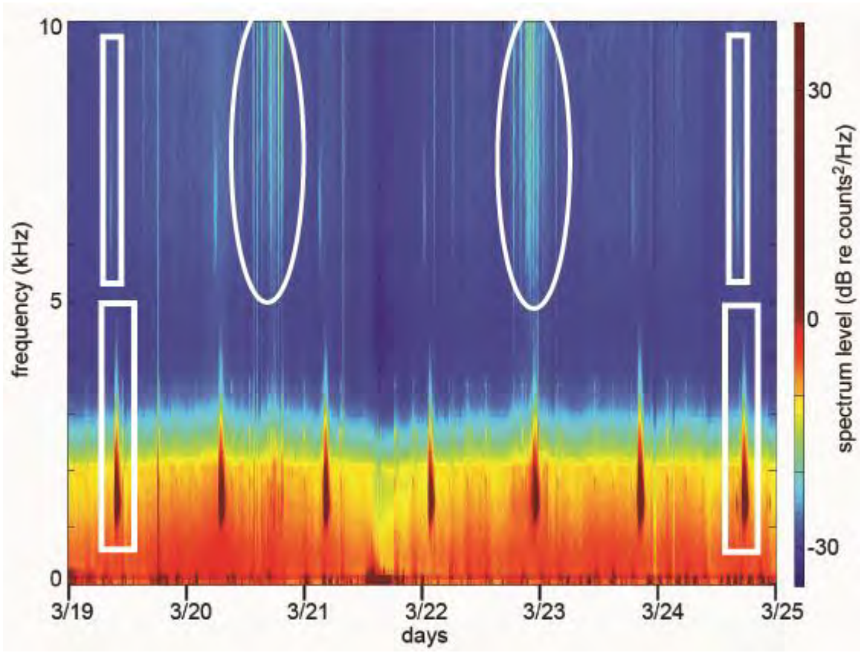

Figure 2 – Long term spectrogram showing a diel pattern at Wake Island. Crepuscular increases in energy were observed between 1 and 5 kHz and in a higher band at Wake Island and many other sites throughout the North Pacific. In this six day plot, the first and last incidences of the noise are highlighted with rectangles. The presence of odontocetes can also be seen strongly on March 20th and 23rd in the signals from 5 kHz to beyond the range of the plot (indicated by ovals). Color online.

The metadata database was used to query sunrise/sunset and lunar illumination/phase at each site over the duration of the deployments. The metadata database connected with the NASA Jet Propulsion Lab Horizons Ephemeris Service (ssd.jpl.nasa.gov/horizons.html) and retrieved data for each of the 35 deployments in a manner that appeared to the user as a normal database query. Illumination was locally processed to derive lunar phase information. This process required about 15 m of real time, which is trivial compared to the amount of work that would have been required had this data been accessed through web interfaces designed for humans. Many areas exhibited diel patterns such as those seen in the 1-6 kHz band ambient noise time series and its demeaned autocorrelation (Fig. 3).

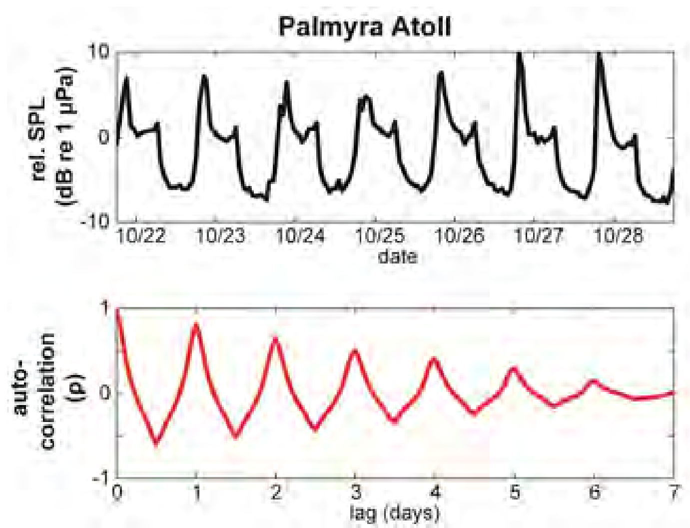

Figure 3 – Ambient noise level in the 1-6 kHz band at Palmyra Atoll (upper plot) and the autocorrelation of the demeaned noise (lower plot). Daily patterns exhibit strong correlation.

Hourly measurements were assigned to one of four categories: dusk, night, dawn, and day. Dusk periods included one hour prior to sunset and three hours thereafter. Similarly, dawn periods included three hours prior to sunrise and the following hour. A nonparametric Kruskal-Wallis test (Zar 2010) showed that for many regions the distributions of the noise in these four time periods were significantly different ($p \leq 0.05$, Fig. 4) at the vast majority of sites.

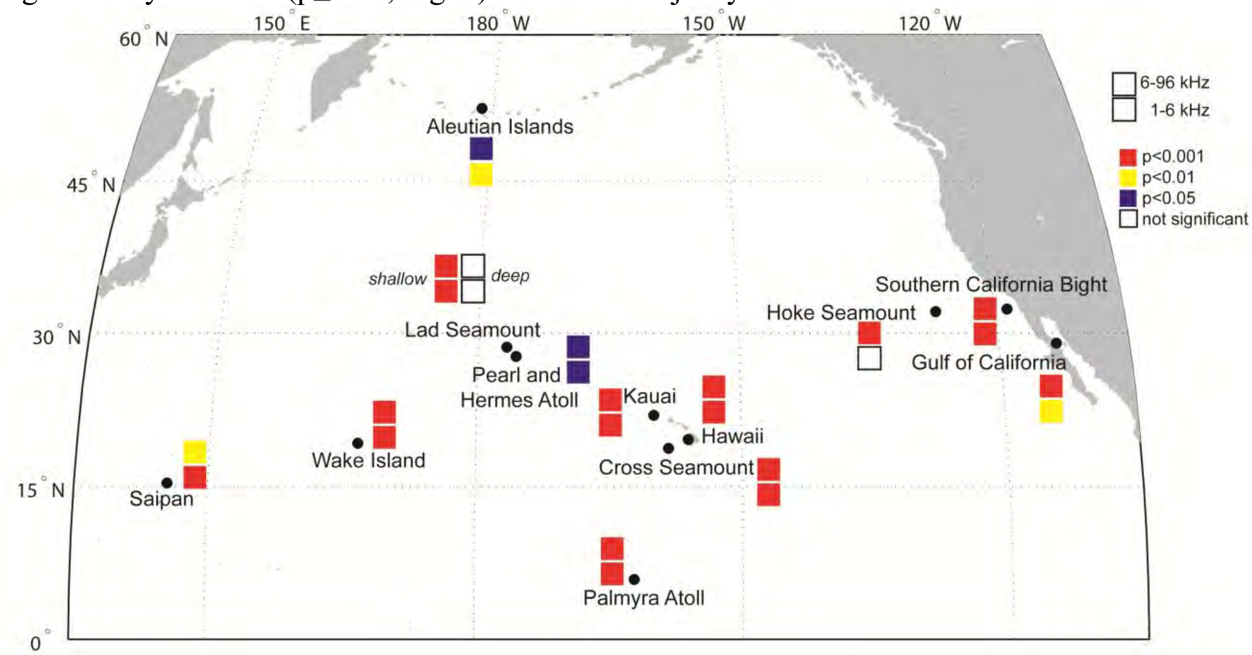

Figure 4 - Evidence of differences in the distributions of dawn, day, dusk, and night ambient noise (nonparametric Kruskal -Wallis test, p-values coded by shading) in two noise bands. The lower box at each site denotes the 1-6 kHz band while the upper box denotes 6-96 kHz band. Lad seamount has deployments at two sites marked shallow and deep. Color online.

The relationship between lunar phase and the noise in different periods was examined (Fig. 5), and a linear-circular correlation statistic (Berens 2009) was computed between each period and lunar phase. Lunar phase was significantly correlated at a number of sites ($p \leq 0.05$, Fig. 6).

The source of the changes in ambient noise is unknown, but with the exception of the ambient noise-lunar correlation in the Aleutian islands which may be due to tidal pull, evidence from other studies suggest that the change in the ambient noise may be due diel activity changes of the mesopelagic boundary layer which has been observed to migrate vertically from deep waters during the day to shallow waters at night. Benoit-Bird et al. (2009) observed an additional horizontal component to this migration in Hawaiian waters with movement between deeper waters and shallower inshore ones. Lammers and Au (unpublished data, personal communication) have also observed similar acoustic patterns and agree with our hypothesis on the origin of the observed noise.

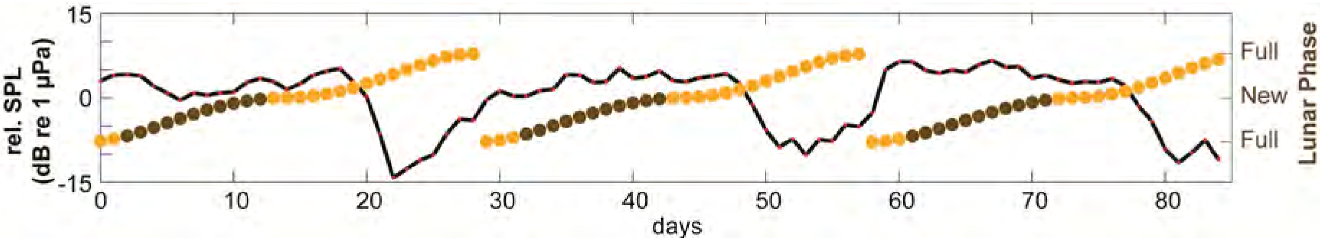

Figure 5 – Coincidence of ambient noise (1-6 kHz band) and lunar phase at Palmyra Atoll. Visibility of moon indicated by light (visible) and dark (obscured) circles.

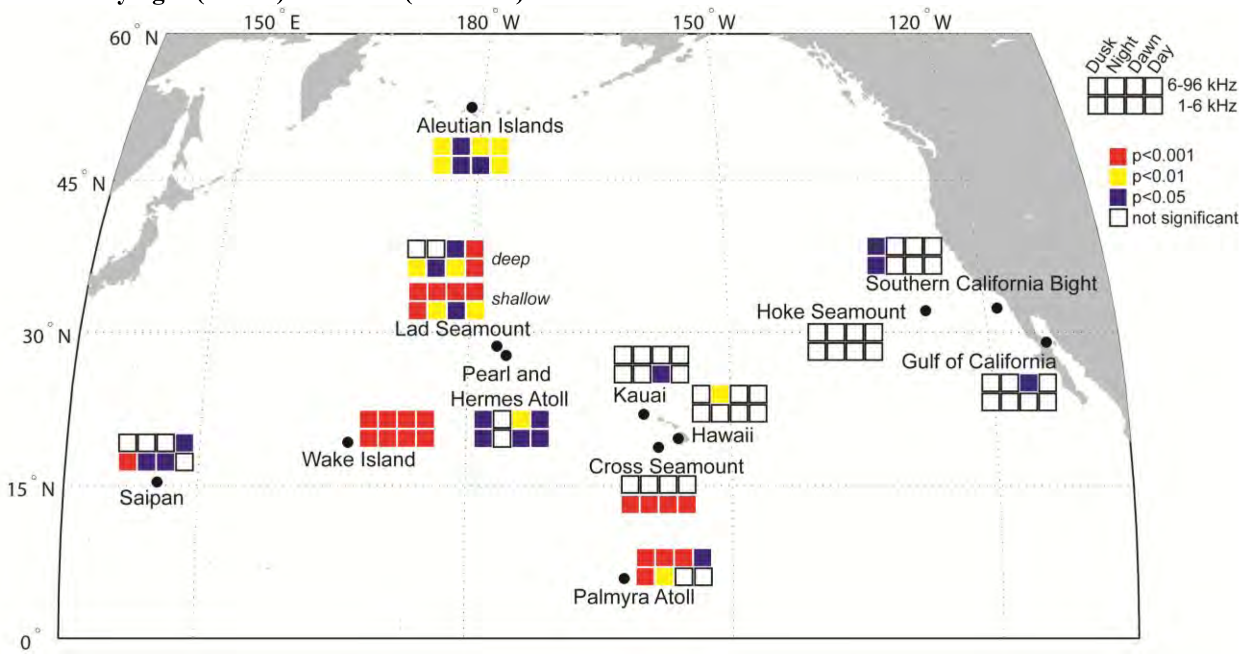

Figure 6 – Significance of linear-circular correlation for ambient noise in different frequency bands (1-6 kHz and 6-96 kHz) and time periods (dusk, night, dawn, and day). Color online.

While these data are preliminary, the ability to detect the arrival of the mesopelagic boundary layer which contains prey species for many cetacean species and possibly judge its abundance based on signal strength, may play a role in the understanding of ecosystem health and habitat use. Longitudinal studies covering vast geographic regions become much more plausible with tools such as the acoustic metadata database that organize and automate access to the data that we wish to use.

IMPACT AND APPLICATIONS

National Security

This effort will enable better understanding of long term trends for cetacean ecosystem use and enable the retention and analysis of long-term time series coupled with oceanographic measurements. Such efforts will help to aid naval mitigation and planning efforts with respect to federally mandated rules on operations in the presence of marine mammals.

Economic Development

Greater understanding of marine mammal habitat use has the potential to enable science-based policy decisions on matters affecting marine mammal habitats such as offshore drilling, wind farms, etc.

Quality of Life

The ability to archive data over long time periods and investigate hypotheses related to oceanographic data will provide opportunities for science-based policy decisions.

Science Education and Communication

This project has the potential to increase the rate of acquisition of knowledge by permitting scientists to change the scope of questions and hypotheses that are posed.

TRANSITIONS

The strong participation by NOAA regional fisheries science centers is one of the strengths of this program and we have been joined this year by Dr. Melissa Soldevilla who will be bringing this work to the NOAA Southeast Fisheries Science Center.

RELATED PROJECTS

OBIS-SEAMAP – Dr. Patrick Halpin (Sloan Foundation, NOPP/National Science Foundation Award NSF-OCE-07-39199)

Marine Geospatial Ecology Tools – Dr. Patrick Halpin (David and Lucile Packard Foundation)

Advanced Detection, Classification, and Localization – Dr. Dave Mellinger (Office of Naval Research Award N00014-11-IP20086_FDOC_24295771)

REFERENCES

Baumann-Pickering, S., A. Sirovic, M.A. Roch, A.E. Simonis, S.M. Wiggins, E.M. Oleson, J.A. Hildebrand (**2011**) Diel and lunar variations of marine ambient sound in the North Pacific (A). *J. Acous. Soc. Am.*, **130**(4), 2536-2536, doi: 10.1121/1.3655131.

Benoit-Bird, K.J., A.D. Dahood, B. Würsig (**2009**) Using active acoustics to compare lunar effects on predator-prey behavior in two marine mammal species. *Mar. Ecol. Prog.-Ser.*, **395**, 119-135, doi: 10.3354/meps07793.

Berens, P. (**2009**) CircStat: A MATLAB Toolbox for Circular Statistics. *J Stat Softw*, **31**(10), 1-21.

Wiggins, S.M., J.A. Hildebrand **(2007**) High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring, Intl. Symp. Underwater Tech., Tokyo, Japan: 551-557, doi: 10.1109/UT.2007.370760

Zar, J.H. (**2010**). *Biostatistical analysis*. Prentice-Hall/Pearson, Upper Saddle River, N.J.: xiii, 944 p. pp.

PUBLICATIONS

The following **presentation** has been completed from this work:

Baumann-Pickering, S., A. Sirovic, M.A. Roch, A.E. Simonis, S.M. Wiggins, E.M. Oleson, J.A. Hildebrand (**2011**) Diel and lunar variations of marine ambient sound in the North Pacific (A). J. Acous. Soc. Am., **130**(4), 2536-2536, doi: 10.1121/1.3655131.